(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: 0 335 521 A1

(12) **EUROPEAN PATENT APPLICATION**

(21) Application number: 89302422.4

(51) Int. Cl.[4]: G10L 3/00 , G10L 9/08

(22) Date of filing: 10.03.89

(30) Priority: 11.03.88 GB 8805795
06.06.88 GB 8813346
24.08.88 GB 8820105

(43) Date of publication of application:
04.10.89 Bulletin 89/40

(84) Designated Contracting States:
AT BE CH DE ES FR GB GR IT LI LU NL SE

(71) Applicant: BRITISH TELECOMMUNICATIONS public limited company
British Telecom Centre, 81 Newgate Street
London EC1A 7AJ(GB)

(72) Inventor: Freeman, Daniel Kenneth
42 Finchley Road Ipswich
Suffolk IP4 2HT(GB)
Inventor: Boyd, Ivan
5 Homefield Capel St Mary
Ipswich Suffolk IP9 2XE(GB)

(74) Representative: Lloyd, Barry George William et al
Intellectual Property Unit British Telecom
Room 1304 151 Gower Street
London WC1E 6BA(GB)

(54) **Voice activity detection.**

(57) Voice activity detector (VAD) for use in an LPC coder in a mobile radio system, uses autocorrelation coefficients R_0, R_1..... of the input signal, weighted and combined, to provide a measure M which depends on the power within that part of the spectrum containing no noise, which is thresholded against a variable threshold to provide a speech/no speech logic output. The measure is

$$M = R_0H_0 + 2\sum_{i=1}^{N} R_iH_i,$$

where H_i are the autocorrelation coefficients of the impulse response of an Nth order FIR inverse noise filter derived from LPC analysis of previous non-speech signal frames. Threshold adaption and coefficient update are controlled by a second VAD responsive to rate of spectral change between frames.

EP 0 335 521 A1

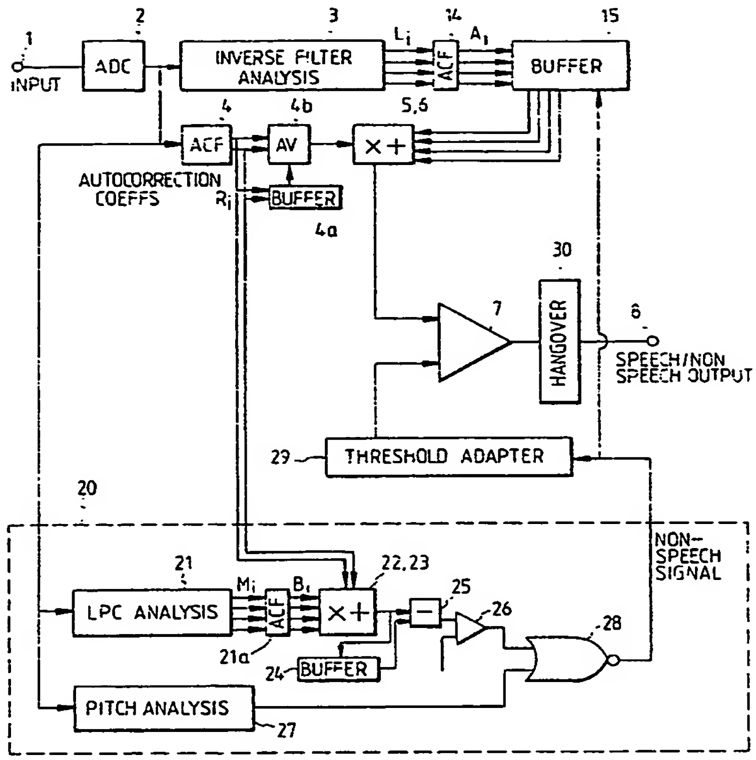

1
INPUT
2
ADC
3
INVERSE FILTER ANALYSIS
L_i
14
ACF
A_i
15
BUFFER
4
ACF
4b
AV
5,6
×+
AUTOCORRECTION COEFFS
R_i
BUFFER
4a
30
HANGOVER
7
6
SPEECH/NON SPEECH OUTPUT
29
THRESHOLD ADAPTER
20
NON-SPEECH SIGNAL
21
M_i
B_i
22,23
LPC ANALYSIS
ACF
×+
25
−
26
28
21a
24
BUFFER
PITCH ANALYSIS
27

FIG. 3.

VOICE ACTIVITY DETECTION

A voice activity detector is a device which is supplied with a signal with the object of detecting periods of speech, or periods containing only noise. Although the present invention is not limited thereto, one application of particular interest for such detectors is in mobile radio telephone systems where the knowledge as to the presence or otherwise of speech can be used exploited by a speech coder to improve the efficient utilisation of radio spectrum, and where also the noise level (from a vehicle-mounted unit) is likely to be high.

The essence of voice activity detection is to locate a measure which differs appreciably between speech and non-speech periods. In apparatus which includes a speech coder, a number of parameters are readily available from one or other stage of the coder, and it is therefore desirable to economise on processing needed by utilising some such parameter. In many environments, the main noise sources occur in known defined areas of the frequency spectrum. For example, in a moving car much of the noise (eg, engine noise) is concentrated in the low frequency regions of the spectrum. Where such knowledge of the spectral position of noise is available, it is desirable to base the decision as to whether speech is present or absent upon measurements taken from that portion of the spectrum which contains relatively little noise. It would, of course, be possible in practice to pre-filter the signal before analysing to detect speech activity, but where the voice activity detector follows the output of a speech coder, prefiltering would distort the voice signal to be coded.

According to a first aspect of the invention there is provided voice activity detection apparatus comprising means for receiving an input signal, means for estimating the noise signal component of the input signal, means for continually forming a measure M of the spectral similarity between a portion of the input signal and the noise signal, and means for comparing a parameter derived from the measure M with a threshold value T to produce an output to indicate the presence or absence of speech in dependence upon whether or not that value is exceeded.

According to a second aspect of the invention there is provided voice activity detection apparatus comprising: means for continually forming a spectral distortion measure of the similarity between a portion of the input signal and earlier portions of the input signal and means for comparing the degree of variation between successive values of the measure with a threshold value to produce an output incating the presence or absence of speech in dependence upon whether or not that value is exceeded.

Preferably, the measure is the Itakura-Saito Distortion Measure.

Other aspects of the present invention are as defined in the claims.

Some embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram of a first embodiment of the invention;

Figure 2 shows a second embodiment of the invention;

Figure 3 shows a third, preferred embodiment of the invention.

The general principle underlying a first Voice Activity Detector according to the a first embodiment of the invention is as follows.

A frame of n signal samples (s_0, s_1, s_2, s_3, s_4 ... s_{n-1}) will, when passed through a notional fourth order finite impulse response (FIR) digital filter of impulse response (1, h_0, h_1, h_2, h_3), result in a filtered signal (ignoring samples from previous frames)

$s' =$

(s_0),

$(s_1 + h_0s_0)$,

$(s_2 + h_0s_1 + h_1s_0)$,

$(s_3 + h_0s_2 + h_1s_1 + h_2s_0)$,

$(s_4 + h_0s_3 + h_1s_2 + h_2s_1 + h_1s_0)$,

$(s_5 + h_0s_4 + h_1s_3 + h_2s_2 + h_3s_1)$,

$(s_6 + h_0s_5 + h_1s_4 + h_2s_3 + h_3s_2)$,

$(s_7 ...)$

The zero order autocorrelation coefficient is the sum of each term squared, which may be normalized i.e. divided by the total number of terms (for constant frame lengths it is easier to omit the division); that of the filtered signal is thus

$$R'_0 = \sum_{i=0}^{n-1} (s'_i)^2$$

and this is therefore a measure of the power of the notional filtered signal s' - in other words, of that part of the signal s which falls within the passband of the notional filter.

Expanding, neglecting the first 4 terms,

$R'_0 = (s_4 + h_0s_3 + h_1s_2 + h_2s_1 + h_3s_0)^2$
$+ (s_5 + h_0s_4 + h_1s_3 + h_2s_2 + h_3s_1)^2$
$+ \ldots$
$= s_4^2 + h_0s_4s_3 + h_1s_4s_2 + h_2s_4s_1 + h_3s_4s_0$
$+ h_0s_4s_3 + h_0^2 s_0^2 + h_0h_1s_3s_2 + h_0h_2s_3s_1 + h_0h_3s_3s_0$
$+ h_1s_4s_2 + h_0h_1s_3s_2 + h_1^2 s_2^2 + h_1h_2s_2s_1 + h_1h_3s_2s_0$
$+ h_2s_4s_1 + h_0h_1s_3s_1 + h_1h_2s_2s_1 + h_2^2 s_1^2 + h_2h_3s_1s_0$
$+ h_3s_4s_0 + h_0h_3s_3s_0 + h_1h_3s_2s_0 + h_2h_3s_1s_0 + h_3^2 s_0^2$
$+ \ldots$
$= R_0 (1 + h_0^2 + h_1^2 + h_2^2 + h_3^2)$
$+ R_1 (2h_0 + 2h_0h_1 + 2h_1h_2 + 2h_2h_3)$
$+ R_2 (2h_1 + 2h_1h_3 + 2h_0h_2)$
$+ R_3 (2h_2 + 2h_0h_3)$
$+ R_4 (2h_3)$

So R'_0 can be obtained from a combination of the autocorrelation coefficients R_i, weighted by the bracketed constants which determine the frequency band to which the value of R'_0 is responsive. In fact, the bracketed terms are the autocorrelation coefficients of the impulse response of the notional filter, so that the expression above may be simplified to

$$R'_0 = R_0H_0 + 2\sum_{i=1}^{N} R_iH_i, \qquad \ldots\ldots\ldots\ldots\ (1)$$

where N is the filter order and H_i are the (un-normalised) autocorrelation coefficients of the impulse response of the filter.

In other words, the effect on the signal autocorrelation coefficients of filtering a signal may be simulated by producing a weighted sum of the autocorrelation coefficients of the (unfiltered) signal, using the impulse response that the required filter would have had.

Thus, a relatively simple algorithm, involving a small number of multiplication operations, may simulate the effect of a digital filter requiring typically a hundred times this number of multiplication operations.

This filtering operation may alternatively be viewed as a form of spectrum comparison, with the signal spectrum being matched against a reference spectrum (the inverse of the response of the notional filter). Since the notional filter in this application is selected so as to approximate the inverse of the noise spectrum, this operation may be viewed as a spectral comparison between speech and noise spectra, and the zeroth autocorrelation coefficient thus generated (i.e. the energy of the inverse filtered signal) as a measure of dissimilarity between the spectra. The Itakura-Saito distortion measure is used in LPC to assess the match between the predictor filter and the input spectrum, and in one form is expressed as

$$M = R_0A_0 + 2\sum_{i=1}^{N} R_iA_i,$$

where A_0 etc are the autocorrelation coefficients of the LPC parameter set. It will be seen that this is closely similar to the relationship derived above, and when it is remembered that the LPC coefficients are the taps of an FIR filter having the inverse spectral response of the input signal so that the LPC coefficient set is the impulse response of the inverse LPC filter, it will be apparent that the Itakura-Saito Distortion Measure is in

fact merely a form of equation 1, wherein the filter response H is the inverse of the spectral shape of an all-pole model of the input signal.

In fact, it is also possible to transpose the spectra, using the LPC coefficients of the test spectrum and the autocorrelation coefficients of the reference spectrum, to obtain a different measure of spectral similarity.

The I-S Distortion measure is further discussed in "Speech Coding based upon Vector Quantisation" by A Buzo, A H Gray, R M Gray and J D Markel, IEEE Trans on ASSP, Vol ASSP-28, No 5, October 1980.

Since the frames of signal have only a finite length, and a number of terms (N, where N is the filter order) are neglected, the above result is an approximation only; it gives, however, a surprisingly good indicator of the presence or absence of speech and thus may be used as a measure M in speech detection. In an environment where the noise spectrum is well known and stationary, it is quite possible to simply employ fixed h_0, h_1 etc coefficients to model the inverse noise filter.

However, apparatus which can adapt to different noise environments is much more widely useful.

Referring to Figure 1, in a first embodiment, a signal from a microphone (not shown) is received at an input 1 and converted to digital samples s at a suitable sampling rate by an analogue to digital converter 2. An LPC analysis unit 3 (in a known type of LPC coder) then derives, for successive frames of n (eg 160) samples, a set of N (eg 8 or 12) LPC filter coefficients L_i which are transmitted to represent the input speech. The speech signal s also enters a correlator unit 4 (normally part of the LPC coder 3 since the autocorrelation vector R_i of the speech is also usually produced as a step in the LPC analysis although it will be appreciated that a separate correlator could be provided). The correlator 4 produces the autocorrelation vector R_i, including the zero order correlation coefficient R_0 and at least 2 further autocorrelation coefficients R_1, R_2, R_3. These are then supplied to a multiplier unit 5.

A second input 11 is connected to a second microphone located distant from the speaker so as to receive only background noise. The input from this microphone is converted to a digital input sample train by AD convertor 12 and LPC analysed by a second LPC analyser 13. The "noise" LPC coefficients produced from analyser 13 are passed to correlator unit 14, and the autocorrelation vector thus produced is multiplied term by term with the autocorrelation coefficients R_i of the input signal from the speech microphone in multiplier 5 and the weighted coefficients thus produced are combined in adder 6 according to Equation 1, so as to apply a filter having the inverse shape of the noise spectrum from the noise-only microphone (which in practice is the same as the shape of the noise spectrum in the signal-plus-noise microphone) and thus filter out most of the noise. The resulting measure M is thresholded by thresholder 7 to produce a logic output 8 indicating the presence or absence of speech; if M is high, speech is deemed to be present.

This embodiment does, however, require two microphones and two LPC analysers, which adds to the expense and complexity of the equipment necessary.

Alternatively, another embodiment uses a corresponding measure formed using the autocorrelations from the noise microphone 11 and the LPC coefficients from the main microphone 1, so that an extra autocorrelator rather than an LPC analyser is necessary.

These embodiments are therefore able to operate within different environments having noise at different frequencies, or within a changing noise spectrum in a given environment.

Referring to Figure 2, in the preferred embodiment of the invention, there is provided a buffer 15 which stores a set of LPC coefficients (or the autocorrelation vector of the set) derived from the microphone input 1 in a period identified as being a "non speech" (ie noise only) period. These coefficients are then used to derive a measure using equation 1, which also of course corresponds to the Itakura-Saito Distortion Measure, except that a single stored frame of LPC coefficients corresponding to an approximation of the inverse noise spectrum is used, rather than the present frame of LPC coefficients.

The LPC coefficient vector L_i output by analyser 3 is also routed to a correlator 14, which produces the autocorrelation vector of the LPC coefficient vector. The buffer memory 15 is controlled by the speech/non-speech output of thresholder 7, in such a way that during "speech" frames the buffer retains the "noise" autocorrelation coefficients, but during "noise" frames a new set of LPC coefficients may be used to update the buffer, for example by a multiple switch 16, via which outputs of the correlator 14, carrying each autocorrelation coefficient, are connected to the buffer 15. It will be appreciated that correlator 14 could be positioned after buffer 15. Further, the speech/no-speech decision for coefficient update need not be from output 8, but could be (and preferably is) otherwise derived.

Since frequent periods without speech occur, the LPC coefficients stored in the buffer are updated from time to time, so that the apparatus is thus capable of tracking changes in the noise spectrum. It will be appreciated that such updating of the buffer may be necessary only occasionally, or may occur only once at the start of operation of the detector, if (as is often the case) the noise spectrum is relatively stationary

over time, but in a mobile radio environment frequent updating is preferred.

In a modification of this embodiment, the system initially employs equation 1 with coefficient terms corresponding to a simple fixed high pass filter, and then subsequently starts to adapt by switching over to using "noise period" LPC coefficients. If, for some reason, speech detection fails, the system may return to using the simple high pass filter.

It is possible to normalise the above measure by dividing through by R_0, so that the expression to be thresholded has the form

$$M = A_0 + 2\sum_{i=1}^{N} \frac{R_i A_i}{R_0}$$

This measure is independent of the total signal energy in a frame and is thus compensated for gross signal level changes, but gives rather less marked contrast between "noise" and "speech" levels and is hence preferably not employed in high-noise environments.

Instead of employing LPC analysis to derive the inverse filter coefficients of the noise signal (from either the noise microphone or noise only periods, as in the various embodiments described above), it is possible to model the inverse noise spectrum using an adaptive filter of known type; as the noise spectrum changes only slowly (as discussed below) a relatively slow coefficient adaption rate common for such filters is acceptable. In one embodiment, which corresponds to Figure 1, LPC analysis unit 13 is simply replaced by an adaptive filter (for example a transversal FIR or lattice filter), connected so as to whiten the noise input by modelling the inverse filter, and its coefficients are supplied as before to autocorrelator 14.

In a second embodiment, corresponding to that of Figure 2, LPC analysis means 3 is replaced by such an adapter filter, and buffer means 15 is omitted, but switch 16 operates to prevent the adaptive filter from adapting its coefficients during speech periods.

A second Voice Activity Detector in accordance with another aspect of the invention will now be described.

From the foregoing, it will be apparent that the LPC coefficient vector is simply the impulse response of an FIR filter which has a response approximating the inverse spectral shape of the input signal. When the Itakura-Saito Distortion Measure between adjacent frames is formed, this is in fact equal to the power of the signal, as filtered by the LPC filter of the previous frame. So if spectra of adjacent frames differ little, a correspondingly small amount of the spectral power of a frame will escape filtering and the measure will be low. Correspondingly, a large interframe spectral difference produces a high Itakura-Saito Distortion Measure, so that the measure reflects the spectral similarity of adjacent frames. In a speech coder, it is desirable to minimise the data rate, so frame length is made as long as possible; in other words, if the frame length is long enough, then a speech signal should show a significant spectral change from frame to frame (if it does not, the coding is redundant). Noise, on the other hand, has a slowly varying spectral shape from frame to frame, and so in a period where speech is absent from the signal then the Itakura-Saito Distortion Measure will correspondingly be low - since applying the inverse LPC filter from the previous frame "filters out" most of the noise power.

Typically, the Itakura-Saito Distortion Measure between adjacent frames of a noisy signal containing intermittent speech is higher during periods of speech than periods of noise; the degree of variation (as illustrated by the standard deviation) is higher, and less intermittently variable.

It is noted that the standard deviation of the standard deviation of M is also a reliable measure; the effect of taking each standard deviation is essentially to smooth the measure.

In this second form of Voice Activity Detector, the measured parameter used to decide whether speech is present is preferably the standard deviation of the Itakura-Saito Distortion Measure, but other measures of variance and other spectral distortion measures (based for example on FFT analysis) could be employed.

It is found advantageous to employ an adaptive threshold in voice activity detection. Such thresholds must not be adjusted during speech periods or the speech signal will be thresholded out. It is accordingly necessary to control the threshold adapter using a speech/non-speech control signal, and it is preferable that this control signal should be independent of the output of the threshold adapter.

The threshold T is adaptively adjusted so as to keep the threshold level just above the level of the measure M when noise only is present. Since the measure will in general vary randomly when noise is present, the threshold is varied by determining an average level over a number of blocks, and setting the threshold at a level proportional to this average. In a noisy environment this is not usually sufficient, however, and so an

assessment of the degree of variation of the parameter over several blocks is also taken into account.

The threshold value T is therefore preferably calculated according to

$T = M' + K.d$

where M' is the average value of the measure over a number of consecutive frames, d is the standard deviation of the measure over those frames, and K is a constant (which may typically be 2).

In practice, it is preferred not to resume adaptation immediately after speech is indicated to be absent, but to wait to ensure the fall is stable (to avoid rapid repeated switching between the adapting and non-adapting states).

Referring to Figure 3, in a preferred embodiment of the invention incorporating the above aspects, an input 1 receives a signal which is sampled and digitised by analogue to digital converter (ADC) 2, and supplied to the input of an inverse filter analyser 3, which in practice is part of a speech coder with which the voice activity detector is to work, and which generates coefficients L_i (typically 8) of a filter corresponding to the inverse of the input signal spectrum. The digitised signal is also supplied to an autocorrelator 4, (which is part of analyser 3) which generates the autocorrelation vector R_i of the input signal (or at least as many low order terms as there are LPC coefficients). Operation of these parts of the apparatus is as described in Figres 1 and 2. Preferably, the autocorrelation coefficients R_i are then averaged over several successive speech frames (typically 5-20 ms long) to improve their reliability. This may be achieved by storing each set of autocorrelations coefficients output by autocorrelator 4 in a buffer 4a, and employing an averager 4b to produce a weighted sum of the current autocorrelation coefficients R_i and those from previous frames stored in and supplied from buffer 4a. The averaged autocorrelation coefficients Ra_i thus derived are supplied to weighting and adding means 5,6 which receives also the autocorrelation vector A_i of stored noise-period inverse filter coefficients L_i from an autocorrelator 14 via buffer 15, and forms from Ra_i and A_i a measure M preferably defined as:

$$M = B_0 + 2\sum_{i=1}^{N} \frac{Ra_i B_i}{R_0}$$

This measure is then thresholded by thresholder 7 against a threshold level, and the logical result provides an indication of the presence or absence of speech at output 8.

In order that the inverse filter coefficients L_i correspond to a fair estimate of the inverse of the noise spectrum, it is desirable to update these coefficients during periods of noise (and, of course, not to update during periods of speech). It is, however, preferable that the speech/non-speech decision on which the updating is based does not depend upon the result of the updating, or else a single wrongly identified frame of signal may result in the voice activity detector subsequently going "out of lock" and wrongly identifying following frames. Preferably, therefore, there is provided a control signal generating circuit 20, effectively a separate voice activity detector, which forms an independent control signal indicating the presence or absence of speech to control inverse filter analyser 3 (or buffer 8) so that the inverse filter autocorrelation coefficients A_i used to form the measure M are only updated during "noise" periods. The control signal generator circuit 20 includes LPC analyser 21 (which again may be part of a speech coder and, specifically, may be performed by analyser 3), which produces a set of LPC coefficients M_i corresponding to the input signal and an autocorrelator 21a (which may be performed by autocorrelator 3a) which derives the autocorrelation coefficients B_i of M_i. If analyser 3 is performed by analyser 3, then $M_i = L_i$ and $B_i = A_i$. These autocorrelation coefficients are then supplied to weighting and adding means 22,23 (equivalent to 5, 6) which receive also the autocorrelation vector R_i of the input signal from autocorrelator 4. A measure of the spectral similarity between the input speech frame and the preceding speech frame is thus calculated; this may be the Itakura-Saito distortion measure between R_i of the present frame and B_i of the preceding frame, as disclosed above, or it may instead be derived by calculating the Itakura - Saito distortion measure for R_i and B_i of the present frame, and subtracting (in subtractor 25) the corresponding measure for the previous frame stored in buffer 24, to generate a spectral difference signal (in either case, the measure is preferably energy-normalised by dividing by R_o). The buffer 24 is then, of course, updated. This spectral difference signal, when thresholded by a thresholder 26 is, as discussed above, an indicator of the presence or absence of speech. We have found, however, that although this measure is excellent for distinguishing noise from unvoiced speech (a task which prior art systems are generally incapable of) it is in general rather less able to distinguish noise from voiced speech. Accordingly, there is preferably further provided within circuit 20 a voiced speech detection circuit comprising a pitch analyser 27 (which in practice may

operate as part of a speech coder, and in particular may measure the long term predictor lag value produced in a multipulse LPC coder). The pitch analyser 27 produces a logic signal which is "true" when voiced speech is detected, and this signal, together with the thresholded measure derived from thresholder 26 (which will generally be "true" when unvoiced speech is present) are supplied to the inputs of a NOR gate 28 to generate a signal which is "false" when speech is present and "true" when noise is present. This signal is supplied to buffer 8 (or to inverse filter analyser 3) so that inverse filter coefficients L_i are only updated during noise periods.

Threshold adapter 29 is also connected to receive the non-speech signal control output of control signal generator circuit 20. The output of the threshold adapter 29 is supplied to thresholder 7. The threshold adapter operates to increment or decrement the threshold in steps which are a proportion of the instant threshold value, until the threshold approximates the noise power level (which may conveniently be derived from, for example, weighting and adding circuits 22, 23). When the input signal is very low, it may be desirable that the threshold is automatically set to a fixed, low, level since at the low signal levels the effect of signal quantisation produced by ADC 2 can produce unreliable results.

There may be further provided "hangover" generating means 30, which operates to measure the duration of indications of speech after thresholder 7 and, when the presence of speech has been indicated for a period in excess of a predetermined time constant, the output is held high for a short "hangover" period. In this way, clipping of the middle of low-level speech bursts is avoided, and appropriate selection of the time constant prevents triggering of the hangover generator 30 by short spikes of noise which are falsely indicated as speech. It will of course be appreciated that all the above functions may be executed by a single suitably programmed digital processing means such as a Digital Signal Processing (DSP) chip, as part of an LPC codec thus implemented (this is the preferred implementation), or as a suitably programmed microcomputer or microcontroller chip with an associated memory device.

Conveniently, as described above, the voice detection apparatus may be implemented as part of an LPC codec. Alternatively, where autocorrelation coefficients of the signal or related measures (partial correlation, or "parcor", coefficients) are transmitted to a distant station the voice detection may take place distantly from the codec.

Claims

1. Voice activity detection apparatus comprising means for receiving an input signal, means for estimating the noise signal component of the input signal, means for continually forming a measure M of the spectral similarity between a portion of the input signal and the noise signal component, and means for comparing a parameter derived from the measure M with a threshold value T to produce an output to indicate the presence or absence of speech in dependence upon whether or not that value is exceeded.

2. Apparatus according to claim 1, in which the noise estimating means comprises means for computing the autocorrelation coefficients A_i of the impulse response of an FIR filter having a response approximating the inverse of the short term spectrum of the noise signal component, and the measure forming means comprises means for computing the autocorrelation coefficients R_i of the signal, means connected to receive R_i and A_i, and to calculate M therefrom, the parameter being the value of M.

3. Apparatus according to claim 2, in which

$$M = R_0A_0 + 2\sum R_iA_i.$$

4. Apparatus according to claim 2, in which

$$M = A_0 + \frac{2\sum R_iA_i}{R_0}.$$

5. Apparatus according to any one of claims 2 to 4, further comprising an input arranged to receive a second signal, similarly subject to noise, from which speech is absent, in which the A_i computing means comprises LPC analysis means for deriving values of A_i from the second signal.

6. Apparatus according to any one of claims 2 to 4, further comprising a buffer connected to store data from which the autocorrelation coefficients A_i of the said filter response may be derived, in which the said filter response is periodically calculated from the signal by LPC analysis means, the apparatus being so connected and controlled that the measure M is calculated using the said stored data, and the said stored data is updated only from periods in which speech is indicated to be absent.

7. Apparatus according to any one of claims 1 to 4 in which the noise estimating means includes an adaptive filter.

8. Apparatus according to any one of claims 2 to 6 characterised in that the means for computing the autocorrelation coefficients of the signal are arranged to do so in dependence upon the autocorrelation coefficients of several successive portions of the signal.

9. Apparatus according to claim 1 in which the measure M is a spectral distortion measure.

10. Apparatus according to claim 9 in which the measure M is the Itakura-Saito Distortion measure.

11. Apparatus according to any one of the preceding claims, further comprising means for adjusting the said predetermined threshold T during periods when speech is indicated to be absent.

12. Apparatus detector according to claim 11, further comprising second voice activity detection means arranged to prevent adjustment of the threshold value when speech is present.

13. Apparatus detector as claimed in claim 11 or claim 12, in which the threshold value T is, when adjusted, adjusted to be equal to the mean of the measure plus a term which is a function of the standard deviation of the measure.

14. Voice activity detection apparatus comprising: means for continually forming a spectral distortion measure of the similarity between a portion of the input signal and earlier portions of the input signal and means for comparing the degree of variation between successive values of the measure with a threshold value to produce an output indicating the presence or absence of speech in dependence upon whether or not that value is exceeded.

15. Apparatus as claimed in claim 14, wherein the degree of variation is measured as the standard deviation of a block of successive values of the measure.

16. Apparatus according to Claim 6 further comprising means for indicating the absence of speech to control the updating of the said stored data, the means for indicating the absence of speech being a second voice activity detection means.

17. Apparatus according to Claim 16 and Claim 13 in which the said second voice activity detection means controls both threshold adaption and data updating.

18. Apparatus according to Claim 13 or Claim 16 or Claim 17 in which said second voice activity detection means is apparatus according to Claim 14 or Claim 15.

19. A method of detecting speech activity in a signal, comprising the steps of comparing the signal spectrum with an estimated noise spectrum, forming a variable measure of the spectral similarity therebetween, and comparing that measure with a threshold.

20. A method of detecting speech activity in a signal, comprising the steps of comparing the signal spectrum with a preceding portion of the signal, forming a variable measure of the spectral similarity therebetween, and comparing the degree of variation between successive values of the measure with a threshold.

21. Voice activity detection apparatus substantially as herein described, with reference to Figure 1 or Figure 2 or Figure 3.

22. Apparatus for encoding speech signals including apparatus according to any preceding claim.

23. Mobile telephone apparatus including apparatus according to any preceding claim.

24. A method of detecting speech substantially as herein described.

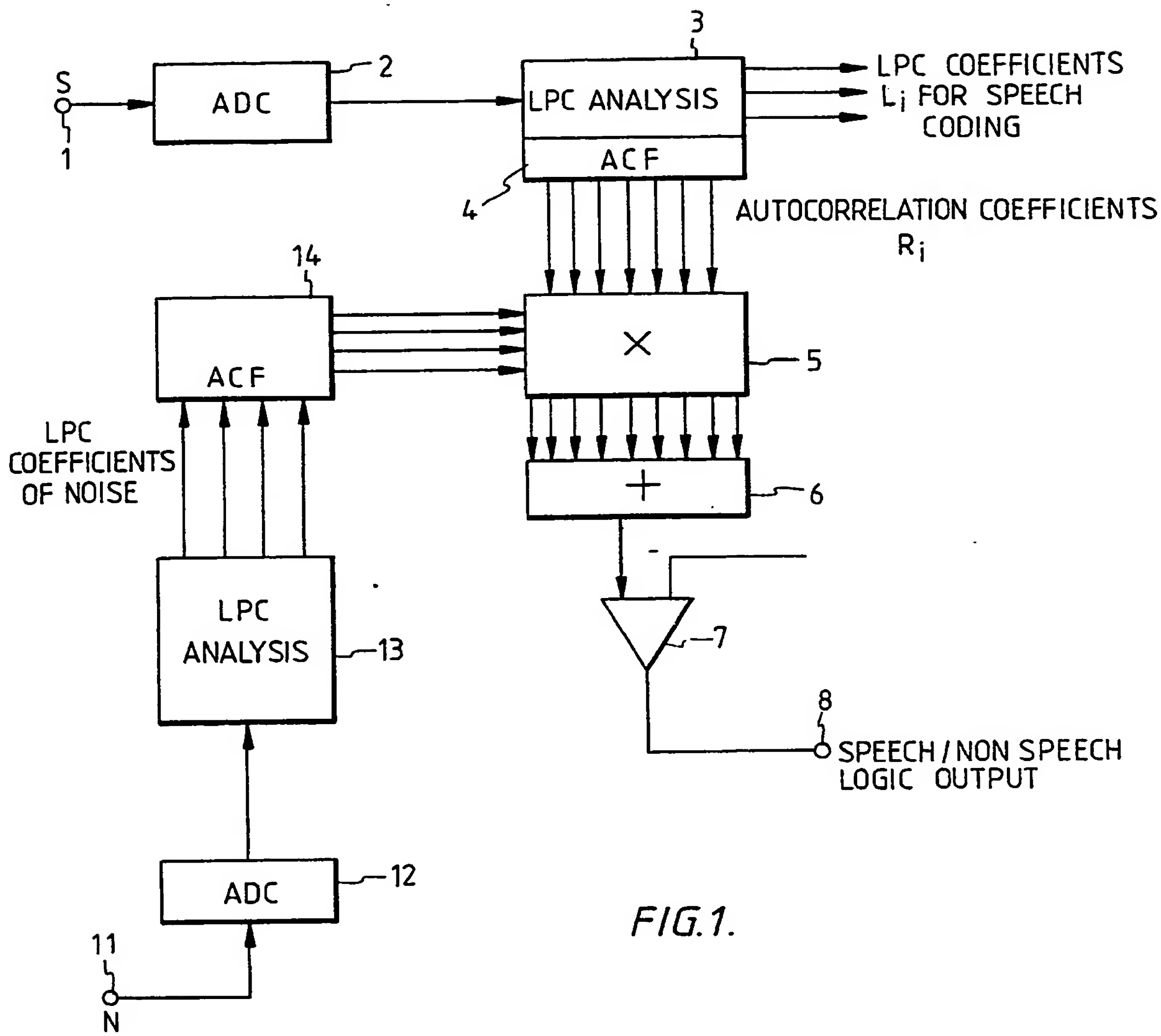

S
1
ADC
2
3
LPC ANALYSIS
ACF
4
LPC COEFFICIENTS
L_i FOR SPEECH
CODING
AUTOCORRELATION COEFFICIENTS
R_i
14
ACF
×
5
LPC
COEFFICIENTS
OF NOISE
+
6
-
7
LPC
ANALYSIS
13
8
SPEECH / NON SPEECH
LOGIC OUTPUT
ADC
12
11
N

FIG.1.

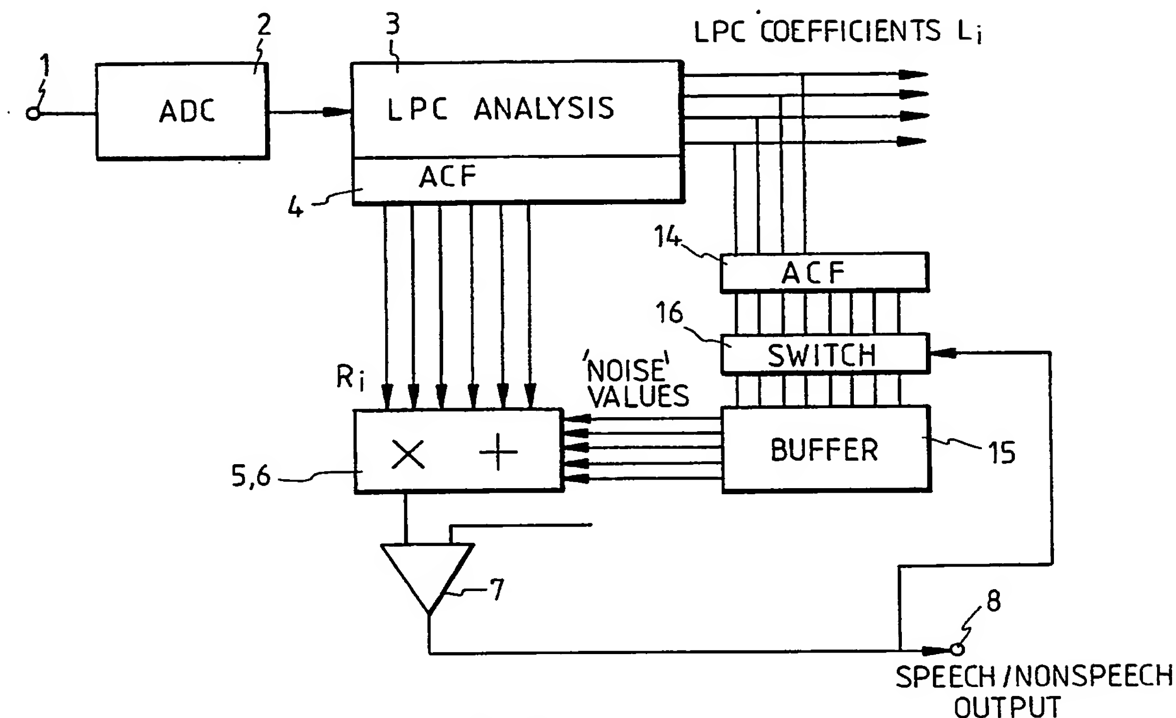

LPC COEFFICIENTS L_i
1
2
ADC
3
LPC ANALYSIS
ACF
4
14
ACF
16
SWITCH
R_i
'NOISE' VALUES
BUFFER
15
5,6
7
8
SPEECH/NONSPEECH OUTPUT

FIG. 2.

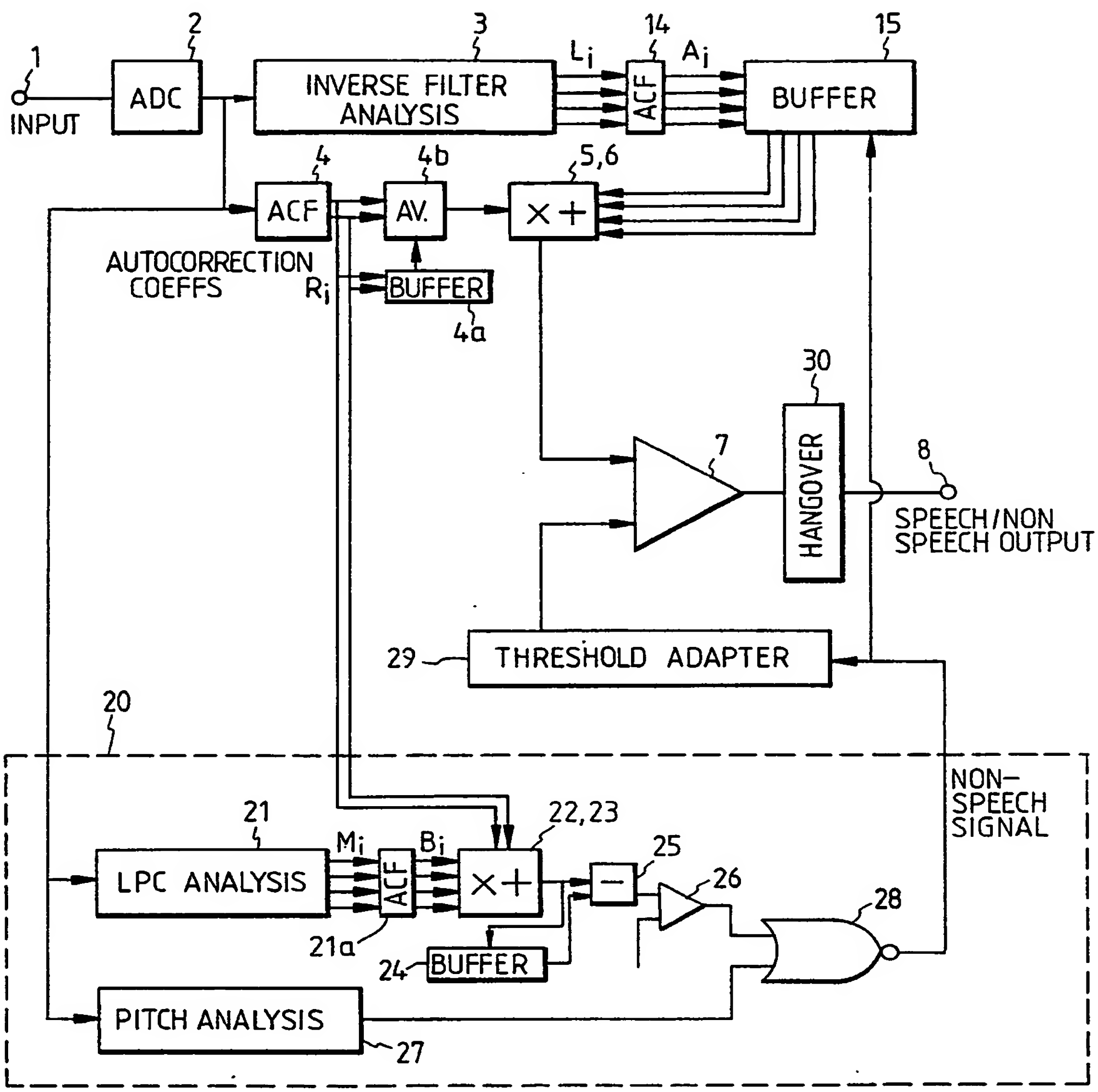

1
INPUT
2
ADC
3
INVERSE FILTER ANALYSIS
L_i
14
ACF
A_i
15
BUFFER
4
ACF
4b
AV.
5,6
× +
AUTOCORRECTION COEFFS
R_i
BUFFER
4a
30
HANGOVER
7
8
SPEECH/NON SPEECH OUTPUT
29
THRESHOLD ADAPTER
20
NON-SPEECH SIGNAL
21
LPC ANALYSIS
M_i
ACF
B_i
× +
22,23
25
−
26
28
21a
24
BUFFER
PITCH ANALYSIS
27

FIG. 3.

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 89 30 2422

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.4)
X	IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, vol. ASSP-25, no. 4, August 1977, pages 338-343, New York, US; L.R. RABINER et al.: "Application of an LPC distance measure to the voiced-unvoiced-silence detection problem" * Page 338, right-hand column, lines 19-23 * ---	1,9,10, 19	G 01 L 3/00 G 10 L 9/08
X	US-A-4 358 738 (L.R. KAHN) * Abstract * ---	1,19	
A	ICASSP'81 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Atlanta, 30th March - 1st April 1981, vol. 3, pages 1082-1085, IEEE, New York, US; C.K. UN et al.: "Improving LPC analysis of noisy speech by autocorrelation subtraction method" * Page 1083, left-hand column, lines 28-33 * ---	6,16	TECHNICAL FIELDS SEARCHED (Int. Cl.4)
A	1977 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, and SIGNAL PROCESSING, Hartford, Connecticut, 9th-11th May 1977, pages 425-428, IEEE, New York, US; R.J. McAULAY: "Optimum speech classification and its application to adaptive noise cancellation" * Abstract * ---	7	G 10 L 3/00 G 10 L 9/08
A	US-A-4 052 568 (J.A. JANKOWSKI) * Abstract * --- -/-	11,12	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	06-07-1989	ARMSPACH J.F.A.M.

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 89 30 2422

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.4)
A	IBM TECHNICAL DISCLOSURE BULLETIN, vol. 22, no. 7, December 1979, pages 2624-2625, New York, US; R.J. JOHNSON et al.: "Speech detector" * Page 2624, lines 1,2; page 2625, lines 8-10 * ---	13,15	
A	EP-A-0 127 718 (COMPAGNIE IBM FRANCE) * Page 10, lines 8-17 * ---	14,20	
A	US-A-4 688 256 (S. YASUNAGA) * Column 3, lines 21-31 * ---	14,20	
A	GB-A-2 061 676 (THE MARCONI CO.) * Abstract * ---	14,20	
A	IEEE TRANSACTIONS ON COMMUNICATIONS, vol. COM-26, no. 1, January 1978, pages 140-145, IEEE, New York, US; P.G. DRAGO et al.: "Digital dynamic speech detectors" * Paragraph 4: "Dynamic speech detector No.1" * ---	14,20	TECHNICAL FIELDS SEARCHED (Int. Cl.4)
A	EP-A-0 178 933 (SHARP K.K.) * Abstract * -----	2	

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	06-07-1989	ARMSPACH J.F.A.M.

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 335 521 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication of patent specification: **24.11.93**

(51) Int. Cl.[5]: **G10L 3/00, G10L 9/08**

(21) Application number: **89302422.4**

(22) Date of filing: **10.03.89**

(54) Voice activity detection.

(30) Priority: **11.03.88 GB 8805795**
06.06.88 GB 8813346
24.08.88 GB 8820105

(43) Date of publication of application:
04.10.89 Bulletin 89/40

(45) Publication of the grant of the patent:
24.11.93 Bulletin 93/47

(84) Designated Contracting States:
AT BE CH DE ES FR GB GR IT LI LU NL SE

(56) References cited:
EP-A- 0 127 718
EP-A- 0 178 933
GB-A- 2 061 676
US-A- 4 052 568
US-A- 4 358 738
US-A- 4 688 256

IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, vol. ASSP-25, no. 4, August 1977, pages 338-343, New York, US; L.R. RABINER et al.: "Application of an LPC distance measure to the voiced-unvoiced-silence detection problem"

(73) Proprietor: **BRITISH TELECOMMUNICATIONS public limited company**
British Telecom Centre,
81 Newgate Street
London EC1A 7AJ(GB)

(72) Inventor: **Freeman, Daniel Kenneth**
42 Finchley Road
Ipswich
Suffolk IP4 2HT(GB)
Inventor: **Boyd, Ivan**
5 Homefield
Capel St Mary
Ipswich Suffolk IP9 2XE(GB)

(74) Representative: **Lloyd, Barry George William et al**
Intellectual Property Unit
British Telecom
Room 1304
151 Gower Street
London WC1E 6BA (GB)

Rank Xerox (UK) Business Services
(3.10/3.6/3.3.1)

ICASSP'81 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, Atlanta, 30th March - 1st April 1981, vol. 3, pages 1082-1085, IEEE, New York, US; C.K. UN et al.: "Improving LPC analysis of noisy speech by autocorrelation subtraction method"

1977 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, and SIGNAL PROCESSING, Hartford, Connecticut, 9th-11th May 1977, pages 425-428, IEEE, New York, US; R.J. McAULAY: "Optimum speech classification and its application to adaptive noise cancellation"

IBM TECHNICAL DISCLOSURE BULLETIN, vol. 22, no. 7, December 1979, pages 2624-2625, New York, US; R.J. JOHNSON et al.: "Speech detector"

IEEE TRANSACTIONS ON COMMUNICATIONS, vol. COM-26, no. 1, January 1978, pages 140-145, IEEE, New York, US; P.G. DRAGO et al.: "Digital dynamic speech detectors"

Description

A voice activity detector is a device which is supplied with a signal with the object of detecting periods of speech, or periods containing only noise. Although the present invention is not limited thereto, one application of particular interest for such detectors is in mobile radio telephone systems where the knowledge as to the presence or otherwise of speech can be used exploited by a speech coder to improve the efficient utilisation of radio spectrum, and where also the noise level (from a vehicle-mounted unit) is likely to be high.

The essence of voice activity detection is to locate a measure which differs appreciably between speech and non-speech periods. In apparatus which includes a speech coder, a number of parameters are readily available from one or other stage of the coder, and it is therefore desirable to economise on processing needed by utilising some such parameter. In many environments, the main noise sources occur in known defined areas of the frequency spectrum. For example, in a moving car much of the noise (eg, engine noise) is concentrated in the low frequency regions of the spectrum. Where such knowledge of the spectral position of noise is available, it is desirable to base the decision as to whether speech is present or absent upon measurements taken from that portion of the spectrum which contains relatively little noise. It would, of course, be possible in practice to pre-filter the signal before analysing to detect speech activity, but where the voice activity detector follows the output of a speech coder, prefiltering would distort the voice signal to be coded.

In US4358738, a voice activity detector is disclosed which compares the input signal with predetermined noise characteristics, by filtering the input signal through a pair of manually balanced bandpass filters (employing analogue components) to form two frequency dependent energy segments. This method is of limited usefulness for many reasons; firstly, such a crude arrangement ignores the fact that many types of noise could have an energy balance between the two bands similar to a speech signal, secondly, balancing the filters is laborious and requires a manual detection of noise periods for balancing, and thirdly, such a device is unable to adjust to changing noise or spectral changes in the environment (or communications channel).

In IEEE transactions on acoustics, speech and signal processing, vol ASSP-25, No. 4, August 1977, page 338-343, Rabiner et al "Application of an LPC distance measure to the voiced unvoiced silence detection problem", there is disclosed a classifier for discriminating between silence, unvoiced speech, and voiced speech which has been transmitted over a telephone line. The method comprises initially using manually classified "silence", "voiced", and "unvoiced" frames of speech signals to drive reference patterns, and then comparing the input signal to each of these using a comparison measure and selecting the reference pattern to which the input signal is closest. This method shares some of the disadvantages of US4358738, in that it requires extensive manual intervention in selecting "silence" frames from training data and forming therefrom the reference pattern, and that since the reference pattern is fixed changes in the environment result in wrong identifications. These problems are greatly exacerbated in high level noise environments (such as a moving vehicle) compared to the low level noise environment (silence over a telephone line) described by Rabiner.

European patent application published as EP-A-0127718 and US patent 4672669 describe a voice activity detection apparatus in which a first test is made on signal amplitude and a second test is based on analysis of changes in the short-term signal spectrum. Specifically, the spectral analysis is performed by comparing the autocorrelation of the signal with that of an earlier portion of the signal deemed to be speech-free.

According to one aspect of the present invention there is provided a voice activity detection apparatus comprising:

(i) means for receiving a first, input, signal;

(ii) means for periodically adaptively generating a second signal representing an estimated noise signal component of the first signal;

(iii) means for periodically forming from the first and second signals a measure of the spectral similarity between a portion of the input signal and the said estimated noise signal component; and

(iv) means for comparing the measure with a threshold value to produce an output indicating the presence or absence of speech;

in which

(v) the generating means includes analysis means operable to produce the coefficients of a filter having a spectral response which is the inverse of the frequency spectrum of one of the said two signals; and

(vi) the measure forming means are operable to produce a measure which is proportional to the zero-order autocorrelation of the other of the said two signals after filtering by a filter having the said

coefficients.

In another aspect, the invention provides a method of detecting voice activity in a first, input, signal, comprising

(a) periodically adaptively generating a second signal representing an estimated noise signal component of the first signal;

(b) periodically forming from the first and second signals a measure of the spectral similarity between a portion of the input signal and the said estimated noise signal component; and

(c) comparing the measure with a threshold value to produce an output indicating the presence or absence of speech;

in which

(d) the generating step includes producing the coefficients of a filter having a spectral response which is the inverse of the frequency spectrum of signals; and

(e) the measure is proportional to the zero-order autocorrelation of the other of the said two signals after filtering by a filter having the said coefficients.

Other aspects of the present invention are as defined in the claims.

Some embodiments of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 is a block diagram of a first embodiment of the invention;

Figure 2 shows a second embodiment of the invention;

Figure 3 shows a third, preferred embodiment of the invention.

The general principle underlying a first Voice Activity Detector according to the a first embodiment of the invention is as follows.

A frame of n signal samples

(s_0, s_1, s_2, s_3, s_4 ... s_{n-1}) will, when passed through a notional fourth order finite impulse response (FIR) digital filter of impulse response (1, h_0, h_1, h_2, h_3), result in a filtered signal (ignoring samples from previous frames)

$s' =$

(s_0),

$(s_1 + h_0 s_0)$,

$(s_2 + h_0 s_1 + h_1 s_0)$,

$(s_3 + h_0 s_2 + h_1 s_1 + h_2 s_0)$,

$(s_4 + h_0 s_3 + h_1 s_2 + h_2 s_1 + h_1 s_0)$,

$(s_5 + h_0 s_4 + h_1 s_3 + h_2 s_2 + h_3 s_1)$,

$(s_6 + h_0 s_5 + h_1 s_4 + h_2 s_3 + h_3 s_2)$,

$(s_7 \ldots)$

The zero order autocorrelation coefficient is the sum of each term squared, which may be normalized i.e. divided by the total number of terms (for constant frame lengths it is easier to omit the division); that of the filtered signal is thus

$$R'_0 = \sum_{i=0}^{n-1} (s'_i)^2$$

and this is therefore a measure of the power of the notional filtered signal s' - in other words, of that part of the signal s which falls within the passband of the notional filter.

Expanding, neglecting the first 4 terms,

$$
\begin{aligned}
R'_0 &= (s_4 + h_0 s_3 + h_1 s_2 + h_2 s_1 + h_3 s_0)^2 \\
&\quad + (s_5 + h_0 s_4 + h_1 s_3 + h_2 s_2 + h_3 s_1)^2 \\
&\quad + \ldots
\end{aligned}
$$

$$
\begin{array}{lllll}
= s_4^2 & + h_0 s_4 s_3 & + h_1 s_4 s_2 & + h_2 s_4 s_1 & + h_3 s_4 s_0 \\
+ h_0 s_4 s_3 & + h_0^2 s_0^2 & + h_0 h_1 s_3 s_2 & + h_0 h_2 s_3 s_1 & + h_0 h_3 s_3 s_0 \\
+ h_1 s_4 s_2 & + h_0 h_1 s_3 s_2 & + h_1^2 s_2^2 & + h_1 h_2 s_2 s_1 & + h_1 h_3 s_2 s_0 \\
+ h_2 s_4 s_1 & + h_0 h_1 s_3 s_1 & + h_1 h_2 s_2 s_1 & + h_2^2 s_1^2 & + h_2 h_3 s_1 s_0 \\
+ h_3 s_4 s_0 & + h_0 h_3 s_3 s_0 & + h_1 h_3 s_2 s_0 & + h_2 h_3 s_1 s_0 & + h_3^2 s_0^2 \\
+ \ldots & & & &
\end{array}
$$

$$
\begin{aligned}
&= R_0 \ (1 + h_0^2 + h_1^2 + h_2^2 + h_3^2) \\
&\quad + R_1 \ (2h_0 + 2h_0 h_1 + 2h_1 h_2 + 2h_2 h_3) \\
&\quad + R_2 \ (2h_1 + 2h_1 h_3 + 2h_0 h_2) \\
&\quad + R_3 \ (2h_2 + 2h_0 h_3) \\
&\quad + R_4 \ (2h_3)
\end{aligned}
$$

So R'_0 can be obtained from a combination of the autocorrelation coefficients R_i, weighted by the bracketed constants which determine the frequency band to which the value of R'_0 is responsive. In fact, the bracketed terms are the autocorrelation coefficients of the impulse response of the notional filter, so that the expression above may be simplified to

$$
R'_0 = R_0 H_0 + 2 \sum_{i=1}^{N} R_i H_i, \qquad \ldots\ldots\ldots\ldots\ldots (1)
$$

where N is the filter order and H_i are the (un-normalised) autocorrelation coefficients of the impulse response of the filter.

In other words, the effect on the signal autocorrelation coefficients of filtering a signal may be simulated by producing a weighted sum of the autocorrelation coefficients of the (unfiltered) signal, using the impulse response that the required filter would have had.

Thus, a relatively simple algorithm, involving a small number of multiplication operations, may simulate the effect of a digital filter requiring typically a hundred times this number of multiplication operations.

This filtering operation may alternatively be viewed as a form of spectrum comparison, with the signal spectrum being matched against a reference spectrum (the inverse of the response of the notional filter).

Since the notional filter in this application is selected so as to approximate the inverse of the noise spectrum, this operation may be viewed as a spectral comparison between speech and noise spectra, and the zeroth autocorrelation coefficient thus generated (i.e. the energy of the inverse filtered signal) as a measure of dissimilarity between the spectra. The Itakura-Saito distortion measure is used in LPC to assess the match between the predictor filter and the input spectrum, and in one form is expressed as

$$M = R_0A_0 + 2\sum_{i=1}^{N} R_iA_i,$$

where A_0 etc are the autocorrelation coefficients of the LPC parameter set. It will be seen that this is closely similar to the relationship derived above, and when it is remembered that the LPC coefficients are the taps of an FIR filter having the inverse spectral response of the input signal so that the LPC coefficient set is the impulse response of the inverse LPC filter, it will be apparent that the Itakura-Saito Distortion Measure is in fact merely a form of equation 1, wherein the filter response H is the inverse of the spectral shape of an all-pole model of the input signal.

In fact, it is also possible to transpose the spectra, using the LPC coefficients of the test spectrum and the autocorrelation coefficients of the reference spectrum, to obtain a different measure of spectral similarity.

The I-S Distortion measure is further discussed in "Speech Coding based upon Vector Quantisation" by A Buzo, A H Gray, R M Gray and J D Markel, IEEE Trans on ASSP, Vol ASSP-28, No 5, October 1980.

Since the frames of signal have only a finite length, and a number of terms (N, where N is the filter order) are neglected, the above result is an approximation only; it gives, however, a surprisingly good indicator of the presence or absence of speech and thus may be used as a measure M in speech detection. In an environment where the noise spectrum is well known and stationary, it is quite possible to simply employ fixed h_0, h_1 etc coefficients to model the inverse noise filter.

However, apparatus which can adapt to different noise environments is much more widely useful.

Referring to Figure 1, in a first embodiment, a signal from a microphone (not shown) is received at an input 1 and converted to digital samples s at a suitable sampling rate by an analogue to digital converter 2. An LPC analysis unit 3 (in a known type of LPC coder) then derives, for successive frames of n (eg 160) samples, a set of N (eg 8 or 12) LPC filter coefficients L_i which are transmitted to represent the input speech. The speech signal s also enters a correlator unit 4 (normally part of the LPC coder 3 since the autocorrelation vector R_i of the speech is also usually produced as a step in the LPC analysis although it will be appreciated that a separate correlator could be provided). The correlator 4 produces the autocorrelation vector R_i, including the zero order correlation coefficient R_0 and at least 2 further autocorrelation coefficients R_1, R_2, R_3. These are then supplied to a multiplier unit 5.

A second input 11 is connected to a second microphone located distant from the speaker so as to receive only background noise. The input from this microphone is converted to a digital input sample train by AD convertor 12 and LPC analysed by a second LPC analyser 13. The "noise" LPC coefficients produced from analyser 13 are passed to correlator unit 14, and the autocorrelation vector thus produced is multiplied term by term with the autocorrelation coefficients R_i of the input signal from the speech microphone in multiplier 5 and the weighted coefficients thus produced are combined in adder 6 according to Equation 1, so as to apply a filter having the inverse shape of the noise spectrum from the noise-only microphone (which in practice is the same as the shape of the noise spectrum in the signal-plus-noise microphone) and thus filter out most of the noise. The resulting measure M is thresholded by thresholder 7 to produce a logic output 8 indicating the presence or absence of speech; if M is high, speech is deemed to be present.

This embodiment does, however, require two microphones and two LPC analysers, which adds to the expense and complexity of the equipment necessary.

Alternatively, another embodiment uses a corresponding measure formed using the autocorrelations from the noise microphone 11 and the LPC coefficients from the main microphone 1, so that an extra autocorrelator rather than an LPC analyser is necessary.

These embodiments are therefore able to operate within different environments having noise at different frequencies, or within a changing noise spectrum in a given environment.

Referring to Figure 2, in the preferred embodiment of the invention, there is provided a buffer 15 which stores a set of LPC coefficients (or the autocorrelation vector of the set) derived from the microphone input

1 in a period identified as being a "non speech" (ie noise only) period. These coefficients are then used to derive a measure using equation 1, which also of course corresponds to the Itakura-Saito Distortion Measure, except that a single stored frame of LPC coefficients corresponding to an approximation of the inverse noise spectrum is used, rather than the present frame of LPC coefficients.

The LPC coefficient vector L_i output by analyser 3 is also routed to a correlator 14, which produces the autocorrelation vector of the LPC coefficient vector. The buffer memory 15 is controlled by the speech/non-speech output of thresholder 7, in such a way that during "speech" frames the buffer retains the "noise" autocorrelation coefficients, but during "noise" frames a new set of LPC coefficients may be used to update the buffer, for example by a multiple switch 16, via which outputs of the correlator 14, carrying each autocorrelation coefficient, are connected to the buffer 15. It will be appreciated that correlator 14 could be positioned after buffer 15. Further, the speech/no-speech decision for coefficient update need not be from output 8, but could be (and preferably is) otherwise derived.

Since frequent periods without speech occur, the LPC coefficients stored in the buffer are updated from time to time, so that the apparatus is thus capable of tracking changes in the noise spectrum. It will be appreciated that such updating of the buffer may be necessary only occasionally, or may occur only once at the start of operation of the detector, if (as is often the case) the noise spectrum is relatively stationary over time, but in a mobile radio environment frequent updating is preferred.

In a modification of this embodiment, the system initially employs equation 1 with coefficient terms corresponding to a simple fixed high pass filter, and then subsequently starts to adapt by switching over to using "noise period" LPC coefficients. If, for some reason, speech detection fails, the system may return to using the simple high pass filter.

It is possible to normalise the above measure by dividing through by R_0, so that the expression to be thresholded has the form

$$M = A_0 + 2\sum_{i=1}^{N} \frac{R_i A_i}{R_0}$$

This measure is independent of the total signal energy in a frame and is thus compensated for gross signal level changes, but gives rather less marked contrast between "noise" and "speech" levels and is hence preferably not employed in high-noise environments.

Instead of employing LPC analysis to derive the inverse filter coefficients of the noise signal (from either the noise microphone or noise only periods, as in the various embodiments described above), it is possible to model the inverse noise spectrum using an adaptive filter of known type; as the noise spectrum changes only slowly (as discussed below) a relatively slow coefficient adaption rate common for such filters is acceptable. In one embodiment, which corresponds to Figure 1, LPC analysis unit 13 is simply replaced by an adaptive filter (for example a transversal FIR or lattice filter), connected so as to whiten the noise input by modelling the inverse filter, and its coefficients are supplied as before to autocorrelator 14.

In a second embodiment, corresponding to that of Figure 2, LPC analysis means 3 is replaced by such an adaptive filter, and buffer means 15 is omitted, but switch 16 operates to prevent the adaptive filter from adapting its coefficients during speech periods.

A second Voice Activity Detector for use with another embodiment of the invention will now be described.

From the foregoing, it will be apparent that the LPC coefficient vector is simply the impulse response of an FIR filter which has a response approximating the inverse spectral shape of the input signal. When the Itakura-Saito Distortion Measure between adjacent frames is formed, this is in fact equal to the power of the signal, as filtered by the LPC filter of the previous frame. So if spectra of adjacent frames differ little, a correspondingly small amount of the spectral power of a frame will escape filtering and the measure will be low. Correspondingly, a large interframe spectral difference produces a high Itakura-Saito Distortion Measure, so that the measure reflects the spectral similarity of adjacent frames. In a speech coder, it is desirable to minimise the data rate, so frame length is made as long as possible; in other words, if the frame length is long enough, then a speech signal should show a significant spectral change from frame to frame (if it does not, the coding is redundant). Noise, on the other hand, has a slowly varying spectral shape from frame to frame, and so in a period where speech is absent from the signal then the Itakura-Saito Distortion Measure will correspondingly be low - since applying the inverse LPC filter from the previous

frame "filters out" most of the noise power.

Typically, the Itakura-Saito Distortion Measure between adjacent frames of a noisy signal containing intermittent speech is higher during periods of speech than periods of noise; the degree of variation (as illustrated by the standard deviation) is also higher, and less intermittently variable.

It is noted that the standard deviation of the standard deviation of M is also a reliable measure; the effect of taking each standard deviation is essentially to smooth the measure.

In this second form of Voice Activity Detector, the measured parameter used to decide whether speech is present is preferably the standard deviation of the Itakura-Saito Distortion Measure, but other measures of variance and other spectral distortion measures (based for example on FFT analysis) could be employed.

It is found advantageous to employ an adaptive threshold in voice activity detection. Such thresholds must not be adjusted during speech periods or the speech signal will be thresholded out. It is accordingly necessary to control the threshold adapter using a speech/non-speech control signal, and it is preferable that this control signal should be independent of the output of the threshold adapter.

The threshold T is adaptively adjusted so as to keep the threshold level just above the level of the measure M when noise only is present. Since the measure will in general vary randomly when noise is present, the threshold is varied by determining an average level over a number of blocks, and setting the threshold at a level proportional to this average. In a noisy environment this is not usually sufficient, however, and so an assessment of the degree of variation of the parameter over several blocks is also taken into account.

The threshold value T is therefore preferably calculated according to

$$T = M' + K.d$$

where M' is the average value of the measure over a number of consecutive frames, d is the standard deviation of the measure over those frames, and K is a constant (which may typically be 2).

In practice, it is preferred not to resume adaptation immediately after speech is indicated to be absent, but to wait to ensure the fall is stable (to avoid rapid repeated switching between the adapting and non-adapting states).

Referring to Figure 3, in a preferred embodiment of the invention incorporating the above aspects, an input 1 receives a signal which is sampled and digitised by analogue to digital converter (ADC) 2, and supplied to the input of an inverse filter analyser 3, which in practice is part of a speech coder with which the voice activity detector is to work, and which generates coefficients L_i (typically 8) of a filter corresponding to the inverse of the input signal spectrum. The digitised signal is also supplied to an autocorrelator 4, (which is part of analyser 3) which generates the autocorrelation vector R_i of the input signal (or at least as many low order terms as there are LPC coefficients). Operation of these parts of the apparatus is as described in Figures 1 and 2. Preferably, the autocorrelation coefficients R_i are then averaged over several successive speech frames (typically 5-20 ms long) to improve their reliability. This may be achieved by storing each set of autocorrelations coefficients output by autocorrelator 4 in a buffer 4a, and employing an averager 4b to produce a weighted sum of the current autocorrelation coefficients R_i and those from previous frames stored in and supplied from buffer 4a. The averaged autocorrelation coefficients Ra_i thus derived are supplied to weighting and adding means 5,6 which receives also the autocorrelation vector A_i of stored noise-period inverse filter coefficients L_i from an autocorrelator 14 via buffer 15, and forms from Ra_i and A_i a measure M preferably defined as:

$$M = A_0 + 2\sum \frac{Ra_i A_i}{R_0},$$

This measure is then thresholded by thresholder 7 against a threshold level, and the logical result provides an indication of the presence or absence of speech at output 8.

In order that the inverse filter coefficients L_i correspond to a fair estimate of the inverse of the noise spectrum, it is desirable to update these coefficients during periods of noise (and, of course, not to update during periods of speech). It is, however, preferable that the speech/non-speech decision on which the updating is based does not depend upon the result of the updating, or else a single wrongly identified frame of signal may result in the voice activity detector subsequently going "out of lock" and wrongly identifying following frames. Preferably, therefore, there is provided a control signal generating circuit 20, effectively a separate voice activity detector, which forms an independent control signal indicating the presence or

absence of speech to control inverse filter analyser 3 (or buffer 8) so that the inverse filter autocorrelation coefficients A_i used to form the measure M are only updated during "noise" periods. The control signal generator circuit 20 includes LPC analyser 21 (which again may be part of a speech coder and, specifically, may be performed by analyser 3), which produces a set of LPC coefficients M_i corresponding to the input signal and an autocorrelator 21a (which may be performed by autocorrelator 3a) which derives the autocorrelation coefficients B_i of M_i. If analyser 21 is performed by analyser 3, then $M_i = L_i$ and $B_i = A_i$. These autocorrelation coefficients are then supplied to weighting and adding means 22,23 (equivalent to 5, 6) which receive also the autocorrelation vector R_i of the input signal from autocorrelator 4. A measure of the spectral similarity between the input speech frame and the preceding speech frame is thus calculated; this may be the Itakura-Saito distortion measure between R_i of the present frame and B_i of the preceding frame, as disclosed above, or it may instead be derived by calculating the Itakura - Saito distortion measure for R_i and B_i of the present frame, and subtracting (in subtractor 25) the corresponding measure for the previous frame stored in buffer 24, to generate a spectral difference signal (in either case, the measure is preferably energy-normalised by dividing by R_o). The buffer 24 is then, of course, updated. This spectral difference signal, when thresholded by a thresholder 26 is, as discussed above, an indicator of the presence or absence of speech. We have found, however, that although this measure is excellent for distinguishing noise from unvoiced speech (a task which prior art systems are generally incapable of) it is in general rather less able to distinguish noise from voiced speech. Accordingly, there is preferably further provided within circuit 20 a voiced speech detection circuit comprising a pitch analyser 27 (which in practice may operate as part of a speech coder, and in particular may measure the long term predictor lag value produced in a multipulse LPC coder). The pitch analyser 27 produces a logic signal which is "true" when voiced speech is detected, and this signal, together with the thresholded measure derived from thresholder 26 (which will generally be "true" when unvoiced speech is present) are supplied to the inputs of a NOR gate 28 to generate a signal which is "false" when speech is present and "true" when noise is present. This signal is supplied to buffer 8 (or to inverse filter analyser 3) so that inverse filter coefficients L_i are only updated during noise periods.

Threshold adapter 29 is also connected to receive the non-speech signal control output of control signal generator circuit 20. The output of the threshold adapter 29 is supplied to thresholder 7. The threshold adapter operates to increment or decrement the threshold in steps which are a proportion of the instant threshold value, until the threshold approximates the noise power level (which may conveniently be derived from, for example, weighting and adding circuits 22, 23). When the input signal is very low, it may be desirable that the threshold is automatically set to a fixed, low, level since at the low signal levels the effect of signal quantisation produced by ADC 2 can produce unreliable results.

There may be further provided "hangover" generating means 30, which operates to measure the duration of indications of speech after thresholder 7 and, when the presence of speech has been indicated for a period in excess of a predetermined time constant, the output is held high for a short "hangover" period. In this way, clipping of the middle of low-level speech bursts is avoided, and appropriate selection of the time constant prevents triggering of the hangover generator 30 by short spikes of noise which are falsely indicated as speech. It will of course be appreciated that all the above functions may be executed by a single suitably programmed digital processing means such as a Digital Signal Processing (DSP) chip, as part of an LPC codec thus implemented (this is the preferred implementation), or as a suitably programmed microcomputer or microcontroller chip with an associated memory device.

Conveniently, as described above, the voice detection apparatus may be implemented as part of an LPC codec. Alternatively, where autocorrelation coefficients of the signal or related measures (partial correlation, or "parcor", coefficients) are transmitted to a distant station the voice detection may take place distantly from the codec.

Claims

1. Voice activity detection apparatus comprising:

   (i) means (1) for receiving a first, input, signal;

   (ii) means (14,15) for periodically adaptively generating a second signal representing an estimated noise signal component of the first signal;

   (iii) means (4,5,6) for periodically forming from the first and second signals a measure M of the spectral similarity between a portion of the input signal and the said estimated noise signal component; and

   (iv) means (7) for comparing the measure M with a threshold value T to produce an output indicating the presence or absence of speech;

characterised in that

(v) the apparatus includes analysis means (13,3) operable to produce the coefficients of a filter having a spectral response which is the inverse of the frequency spectrum of one of the said two signals; and

(vi) the measure forming means (4,5,6) are operable to produce a measure M which is proportional to the zero-order autocorrelation (R'_0) of a signal obtained by filtering of the other of the said two signals by a filter having the said coefficients.

2. Apparatus according to claim 1 in which the analysis means (13,3) includes an adaptive filter.

3. Apparatus according to claim 1, in which the generating means (14,15) are operable to compute the autocorrelation coefficients A_i of the impulse response of the said coefficients and the measure forming means (4) comprises means for computing the autocorrelation coefficients R_i of the said other signal, and means (5,6) connected to receive R_i and A_i, and to calculate the measure M therefrom.

4. Apparatus according to claim 2 in which the means (4) for computing the autocorrelation coefficients R_i of the said other signal are arranged (4a,4b) to do so in dependence upon the autocorrelation coefficients of several successive portions of the signal.

5. Apparatus according to claim 3 or 4, in which

$$M = R_0 A_0 + 2\Sigma R_i A_i.$$

where Ai represents the ith autocorrelation coefficient of the impulse response of said filter.

6. Apparatus according to claim 3 or 4, in which

$$M = A_0 + 2\sum \frac{R_i A_i}{R_0}$$

where A_i represents the i^{th} autocorrelation coefficient of the impulse response of said filter.

7. Apparatus according to any one of claims 1 to 6, in which the said one signal is the second, noise representing, signal and the said other signal is the first, input signal.

8. Apparatus according to claim 7, further comprising an input (11) arranged to receive a second input signal, similarly subject to noise, from which speech is absent, in which the generating means comprise LPC analysis means (13) for deriving values of A_i from the second input signal.

9. Apparatus according to any one of claims 1 to 7, further comprising a buffer (15) connected to store data from which the autocorrelation coefficients A_i of the said filter response may be obtained, in which the said filter response is periodically calculated from the signal by LPC analysis means (3), the apparatus being so connected and controlled that the measure M is calculated using the said stored data, and the said stored data is updated only from periods in which speech is indicated to be absent.

10. Apparatus according to claim 9 further comprising means (20) for indicating the absence of speech to control the updating of the stored data, the means (20) for indicating the absence of speech being a second voice activity detection means (20).

11. Apparatus according to any one of the preceding claims, further comprising means (29) for adjusting the said threshold value T during periods when speech is indicated to be absent.

12. Apparatus according to claim 11, further comprising second voice activity detection means (20) arranged to prevent adjustment of the threshold value when speech is present.

13. Apparatus according to claim 10 further comprising means (20) for adjusting the said threshold value T during periods when speech is indicated to be absent, the said second voice activity detection means (20) being arranged to prevent adjustment of the threshold value when speech is present.

14. Apparatus according to claim 11, 12 or 13 in which the threshold value T is, when adjusted, adjusted to be equal to the mean of the measure plus a term which is a fraction of the standard deviation of the measure.

15. Apparatus according to claim 10, 13 or 14 in which said second voice activity detection means (20) comprises means (4, 21, 21a, 22, 23, 24, 25, 26) for generating a measure of the spectral similarity between a portion of the input signal and earlier portions of the input signal.

16. Apparatus according to claim 15 in which the similarity measure generating means comprises means (4, 21, 22, 23) for providing, from LPC filter data and autocorrelation data relating to a present portion of the input signal, a present distortion measure; means (24) for providing an equivalent past frame distortion measure corresponding to a preceding portion of the input signal, and means (25, 26) for generating a signal indicating the degree of similarity therebetween as an indicator of speech presence or absence.

17. Apparatus according to claim 15 or 16, in which said second voice activity detection means (20) further comprises voiced speech detection means (27) comprising pitch analysis means (27), for generating a signal indicative of the presence of voiced speech, upon which the output of said second voice activity detection means (20) also depends.

18. A method of detecting voice activity in a first, input, signal, comprising

(a) periodically adaptively generating a second signal representing an estimated noise signal component of the first signal;

(b) periodically forming from the first and second signals a measure M of the spectral similarity between a portion of the input signal and the said estimated noise signal component; and

(c) comparing the measure M with a threshold value T to produce an output indicating the presence or absence of speech;

characterised by

(d) the step of producing the coefficients of a filter having a spectral response which is the inverse of the frequency spectrum of one of the said two signals; and in that

(e) the measure M is proportional to the zero-order autocorrelation R'_0 of a signal obtained by filtering of the other of the said two signals by a filter having the said coefficients.

19. Apparatus for encoding speech signals including apparatus according to any one of claims 1 to 17.

20. Mobile telephone apparatus including apparatus according to any one claims 1 to 17.

Patentansprüche

1. Vorrichtung zum Erfassen der Anwesenheit von Sprache, die aufweist:

(i) Eine Einrichtung (1) zum Empfangen eines ersten Eingangssignales;

(ii) eine Einrichtung (14, 15) zum periodischen adaptiven Erzeugen eines zweiten Signales, das eine geschätzte Rauschsignalkomponente des ersten Signales darstellt;

(iii) eine Einrichtung (4, 5, 6) zum periodischen Bilden aus dem ersten und zweiten Signal eines Maßes M der spektralen Ähnlichkeit zwischen einem Abschnitt des Eingangssignales und der geschätzten Rauschsignalkomponente; und

(iv) eine Einrichtung (7) zum Vergleichen des Maßes M mit einem Schwellwert T, um eine Ausgabe zu erzeugen, die die Anwesenheit oder Abwesenheit von Sprache anzeigt;

dadurch gekennzeichnet, daß

(v) die Vorrichtung eine Analyseeinrichtung (13, 3) aufweist, die betreibbar ist, um die Koeffizienten eines Filters, das eine Spektralantwort hat, die die Inverse des Frequenzspektrums eines der beiden Signale ist, zu erzeugen; und

(vi) die maßbildende Einrichtung (4, 5, 6), die betreibbar ist, um ein Maß M zu erzeugen, das proportional zu der Autokorrelation R'_0 nullter Ordnung eines Signales ist, das durch Filtern des

anderen der beiden Signale durch ein Filter erhalten wird, das die Koeffizienten hat.

2. Vorrichtung gemäß Anspruch 1, in der die Analyseeinrichtung (13, 3) ein adaptives Filter aufweist.

3. Vorrichtung gemäß Anspruch 1, in der die erzeugende Einrichtung (14, 15) betreibbar ist, um die Autokorrelationskoeffizienten A_i der Impulsantwort der Koeffizienten zu berechnen, und in der die maßbildende Einheit (4) eine Einrichtung zum Berechnen der Autokorrelationskoeffizienten R_i des anderen Signales aufweist, und eine Einrichtung (5, 6), die verbunden ist, um R_i und A_i zu empfangen und das Maß daraus zu berechnen.

4. Vorrichtung gemäß Anspruch 2, bei der die Einrichtung (4) zum Berechnen der Autokorrelationskoeffizienten R_i des anderen Signales angeordnet ist (4a, 4b), um dies in Abhängigkeit von den Autokorrelationskoeffizienten mehrerer aufeinanderfolgender Abschnitte des Signales zu machen.

5. Vorrichtung gemäß Anspruch 3 oder 4, bei der gilt:

$$M = R_0 A_0 + 2\Sigma\, R_i A_i$$

wobei A_i den i-ten Autokorrelationskoeffizienten der Impulsantwort des Filters darstellt.

6. Vorrichtung gemäß Anspruch 3 oder 4, bei der gilt:

$$M = A_0 + 2\sum \frac{R_i A_i}{R_0} \quad ,$$

wobei A_i den i-ten Autokorrelationskoeffizienten der Impulsantwort des Filters darstellt.

7. Vorrichtung gemäß einem der Ansprüche 1 bis 6, bei der das eine Signal das zweite Rauschen darstellende Signal ist und das andere Signal das erste Eingangssignal ist.

8. Vorrichtung gemäß Anspruch 7, die weiterhin einen Eingang (11) aufweist, der angeordnet ist, um ein zweites Eingangssignal zu empfangen, das ähnlich Rauschen unterworfen ist, von dem Sprache abwesend ist, in dem die erzeugende Einrichtung eine LPC-Analyseeinrichtung (13) aufweist, zum Ableiten der Werte von A_i aus dem zweiten Eingangssignal.

9. Vorrichtung gemäß einem der Ansprüche 1 bis 7, die weiterhin einen Puffer (15) aufweist, der verbunden ist, um Daten zu speichern, aus denen die Autokorrelationskoeffizienten A_i der Filterantwort erhalten werden können, in der die Filterantwort periodisch von dem Signal durch eine LPC-Analyseeinrichtung (3) berechnet wird, wobei die Vorrichtung so verbunden und gesteuert ist, daß das Maß M berechnet wird unter Verwendung der gespeicherten Daten, und wobei die gespeicherten Daten nur von Perioden aktualisiert werden, in denen Sprache als anwesend angezeigt ist.

10. Vorrichtung gemäß Anspruch 9, die weiterhin eine Einrichtung (20) zum Anzeigen der Abwesenheit von Sprache aufweist, um das Aktualisieren der gespeicherten Daten zu steuern, wobei die Einrichtung (20) zum Anzeigen der Abwesenheit von Sprache eine zweite Sprachaktivitätserfassungseinrichtung (20) ist.

11. Vorrichtung gemäß einem der vorhergehenden Ansprüche, die weiterhin eine Einrichtung (29) zum Einstellen des Schwellwertes T während Perioden, wenn Sprache als abwesend angezeigt ist, aufweist.

12. Vorrichtung gemäß Anspruch 11, die weiterhin eine zweite Erfassungseinrichtung (20) für die Anwesenheit von Sprache aufweist, die angeordnet ist, um die Einstellung des Schwellwertes zu verhindern, wenn Sprache vorliegt.

13. Vorrichtung gemäß Anspruch 10, die weiterhin eine Einrichtung (20) zum Einstellen des Schwellwertes T während Perioden aufweist, bei denen Sprache als anwesend angezeigt wird, wobei die zweite

Erfassungseinrichtung (20) für die Anwesenheit von Sprache angeordnet ist, um eine Einstellung des Schwellenwertes zu verhindern, wenn Sprache vorliegt.

14. Vorrichtung gemäß den Ansprüchen 11, 12 oder 13, bei der der Schwellwert T, wenn eingestellt, eingestellt ist, um gleich dem Mittel des Maßes plus einem Term zu sein, der ein Bruchteil der Standardabweichung des Maßes ist.

15. Vorrichtung gemäß Anspruch 10, 13 oder 14, bei dem die zweite Sprachaktivitätserfassungseinrichtung (20) eine Einrichtung (4, 21, 21a, 22, 23, 24, 25, 26) zum Erzeugen eines Maßes der spektralen Ähnlichkeit zwischen einem Abschnitt des Eingabesignales und früherer Abschnitte des Eingabesignales aufweist.

16. Vorrichtung gemäß Anspruch 15, bei der die das Ähnlichkeitsmaß erzeugende Einrichtung Einrichtungen (4, 21, 22, 23) aufweist zum Bereitstellen aus LPC-Filterdaten und Autokorrelationsdaten, die sich auf einen vorliegenden Abschnitt des Eingangssignales beziehen, eines vorliegenden Verzerrungsmaßes, eine Einrichtung (24) zum Bereitstellen eines äquivalenten Verzerrungsmaßes des vergangenen Rahmens, entsprechend einem vorhergehenden Abschnitt des Eingangssignales, und Einrichtungen (25, 26) zum Erzeugen eines Signales, das den Grad der Ähnlichkeit zwischen ihnen als ein Indikator von Sprachanwesenheit oder -abwesenheit anzeigt.

17. Vorrichtung gemäß Anspruch 15 oder 16, bei der die zweite Erfassungseinrichtung (20) für die Anwesenheit von Sprache weiterhin eine Erfassungseinrichtung für stimmhafte Sprache (27) aufweist, die eine Tonhöheanalyseeinrichtung (27) aufweist zum Erzeugen eines Signales, das die Anwesenheit von stimmhafter Sprache anzeigt, von dessen Ausgabe die zweite Erfassungseinrichtung (20) für die Anwesenheit von Sprache ebenfalls abhängt.

18. Verfahren zum Erfassen der Anwesenheit von Sprache in einem ersten Eingangssignal, das aufweist:
(a) Periodisches adaptives Erzeugen eines zweiten Signales, das eine geschätzte Rauschsignalkomponente des ersten Signales darstellt;
(b) periodisches Bilden aus dem ersten und zweiten Signal eines Maßes M der spektralen Ähnlichkeit zwischen einem Abschnitt des Eingangssignales und der geschätzten Rauschsignalkomponente; und
(c) Vergleichen des Maßes M mit einem Schwellwert T, um eine Ausgabe zu produzieren, die die Anwesenheit oder Abwesenheit von Sprache anzeigt;
dadurch gekennzeichnet, daß
(d) der Schritt des Produzierens der Koeffizienten eines Filters, das eine Spektralantwort hat, die die Inverse des Frequenzspektrums eines der beiden Signale ist; und darin, daß
(e) das Maß M proportional zu der Autokorrelation R'_0 nullter Ordnung eines Signales ist, das durch Filtern des anderen der beiden Signale durch ein Filter erhalten wird, der die Koeffizienten hat.

19. Vorrichtung zum Codieren von Sprachsignalen, die eine Vorrichtung gemäß einem der Ansprüche 1 bis 17 aufweist.

20. Mobiltelefonvorrichtung, die eine Vorrichtung gemäß einem der Ansprüche 1 bis 17 aufweist.

Revendications

1. Appareil de détection d'activité vocale comprenant:
(i) un moyen de réception (1) d'un premier signal d'entrée;
(ii) un moyen de génération périodique adaptative (14, 15) d'un deuxième signal représentant une composante estimée de signal de bruit du premier signal;
(iii) un moyen de formation périodique (4, 5, 6), à partir du premier et du deuxième signaux, d'une mesure M de la similitude spectrale entre une partie du signal d'entrée et ladite composante estimée de signal de bruit; et
(iv) un moyen de comparaison (7) de la mesure M avec une valeur de seuil T afin de produire une sortie indiquant la présence ou l'absence de parole;
caractérisé en ce que:

(v) l'appareil inclut un moyen d'analyse (13, 3) qui peut être mis en oeuvre pour produire les coefficients d'un filtre dont la réponse spectrale est l'inverse du spectre de fréquence d'un premier desdits deux signaux;

(vi) les moyens de formation (4, 5, 6) de mesure peuvent être mis en oeuvre pour produire une mesure M qui est proportionnelle à l'autocorrélation d'ordre zéro (R'_0) d'un signal obtenu en filtrant, au moyen d'un filtre possédant lesdits coefficients, l'autre desdits deux signaux.

2. Appareil selon la revendication 1 dans lequel le moyen d'analyse (13, 3) inclut un filtre adaptatif.

3. Appareil selon la revendication 1, dans lequel les moyens générateurs (14, 15) peuvent être mis en oeuvre pour calculer les coefficients A_i d'autocorrélation A_i de la réponse d'impulsion desdits coefficients et le moyen de formation (4) de mesure comprend un moyen de calcul des coefficients d'autocorrélation (R_i) dudit autre signal, et les moyens (5, 6) sont reliés de manière à recevoir R_i etA_i et à calculer à partir de ceux-ci la mesure M.

4. Appareil selon la revendication 2 dans lequel le moyen de calcul (4) des coefficients d'autocorrélation R_i dudit autre signal sont agencés (4a, 4b) de manière à les calculer en fonction des coefficients d'autocorrélation de plusieurs autres parties du signal.

5. Appareil selon la revendication 3 ou 4 dans lequel

$M = R_0A_0 + 2 \Sigma R_iA_i$

où Ai représente le i-ième coefficient d'autocorrélation de la réponse d'impulsion dudit filtre.

6. Appareil selon la revendication 3 ou 4, dans lequel

$$M = R_0A_0 + 2 \Sigma \frac{R_iA_i}{R_0}$$

où Ai représente le i-ième coefficient d'autocorrélation de la réponse d'impulsion dudit filtre.

7. Appareil selon l'une quelconque des revendications 1 à 6 dans lequel ledit premier signal est le deuxième signal, représentant le bruit, et ledit autre signal est le premier signal, d'entrée.

8. Appareil selon la revendication 7, comprenant en outre une entrée (11) agencée de manière à recevoir un deuxième signal d'entrée, sujet lui aussi à un bruit, dont une parole est absente, dans lequel le moyen générateur comprend un moyen d'analyse (13) à codage à prédiction linéaire, ou LPC, pour dériver des valeurs de A_i à partir du deuxième signal d'entrée.

9. Appareil selon l'une quelconque des revendications 1 à 7, comprenant en outre un tampon (15) relié de manière à mémoriser des données à partir desquelles peuvent être obtenus les coefficients d'autocorrélation A_i de ladite réponse de filtre, dans lequel ladite réponse de filtre est calculée périodiquement à partir du signal par un moyen d'analyse (3) de signal à codage à prédiction linéaire, l'appareil étant relié et commandé d'une manière telle que la mesure M est calculée en utilisant ladite donnée mémorisée, et la donnée mémorisée n'est mise à jour qu'à partir de périodes dans lesquelles la parole est indiquée comme absente.

10. Appareil selon la revendication 9, comprenant en outre un moyen (20) d'indication de l'absence de parole pour commander la mise à jour de la donnée mémorisée, le moyen (20) d'indication de l'absence de parole étant un deuxième moyen de détection (20) d'activité vocale.

11. Appareil selon l'une quelconque des revendications précédentes comprenant en outre un deuxième moyen (29 d'ajustement de ladite valeur de seuil T pendant des périodes dans lesquelles la parole est indiquée comme absente.

12. Appareil selon la revendication 11, comprenant en outre un deuxième moyen de détection (20) d'activité vocale, agencé de manière à empêcher un ajustement de la valeur de seuil lorsqu'une parole est présente.

13. Appareil selon la revendication 10 comprenant en outre un moyen (20) d'ajustement (20) de ladite valeur de seuil T pendant des périodes où il est indiqué qu'une parole est absente, ledit deuxième moyen (20) de détection d'activité vocale étant agencé de façon à empêcher un ajustement de la valeur de seuil lorsqu'une parole est présente.

14. Appareil selon la revendication 11, 12 ou 13, dans lequel la valeur de seuil T est ajustée, lorsque elle l'est, de manière à être égale à la moyenne de la mesure, augmentée d'un terme qui est une fraction de l'écart type de la mesure.

15. Appareil selon la revendication 10, 13 ou 14 dans lequel ledit deuxième moyen (20) de détection d'activité vocale comprend un moyen (4, 21, 21a, 22, 23, 24, 25, 26) de génération d'une mesure de similitude spectrale entre une partie du signal d'entrée et des parties antérieures du signal d'entrée.

16. Appareil selon la revendication 15 dans lequel le moyen générateur de mesure de similitude comprend des moyens (4, 21, 22, 23) de production d'une mesure actuelle de distorsion, à partir de données de filtre de codage à prédiction linéaire et de données d'autocorrélation concernant une partie actuelle du signal d'entrée; un moyen (24) de production d'une mesure équivalente de distorsion de structure passée, correspondant à une partie précédente du signal d'entrée; et des moyens (25, 26) de génération d'un signal indiquant le degré de similitude entre ces mesures tant qu'indicateur de la présence ou de l'absence d'une parole.

17. Appareil selon la revendication 15 ou 16, dans lequel ledit deuxième moyen de détection (20) d'activité vocale comprend en outre un moyen de détection (27) de parole voisée comprenant un moyen d'analyse de hauteur sonore afin d'engendrer un signal indicatif de la présence d'une parole voisée, dont dépend aussi la sortie du deuxième moyen de détection (20) d'activité vocale.

18. Un procédé de détection d'activité vocale dans un premier signal, d'entrée, comprenant les étapes consistant à:

(a) engendrer périodiquement de façon adaptative un deuxième signal représentant une composante estimée d'un signal de bruit du premier signal;

(b) former périodiquement, à partir du premier et du deuxième signaux, une mesure M de la similitude spectrale entre une partie du signal d'entrée et ladite composante estimée de signal de bruit; et

(c) comparer la mesure M à une valeur de seuil T afin de produire une sortie indiquant la présence ou l'absence d'une parole;

caractérisé par

(d) l'étape de production des coefficients d'un filtre dont la réponse spectrale est l'inverse du spectre de fréquence d'un premier desdits deux signaux; et par le fait que

la mesure M est proportionnelle à l'autocorrélation R'_0 d'ordre zéro d'un signal obtenu en filtrant l'autre desdits deux signaux par un filtre possédant lesdits coefficients.

19. Appareil d'encodage de signaux de parole incluant l'appareil selon l'une quelconque des revendications 1 à 17.

20. Appareil téléphonique mobile incluant un appareil selon l'une quelconque des revendications 1 à 17.

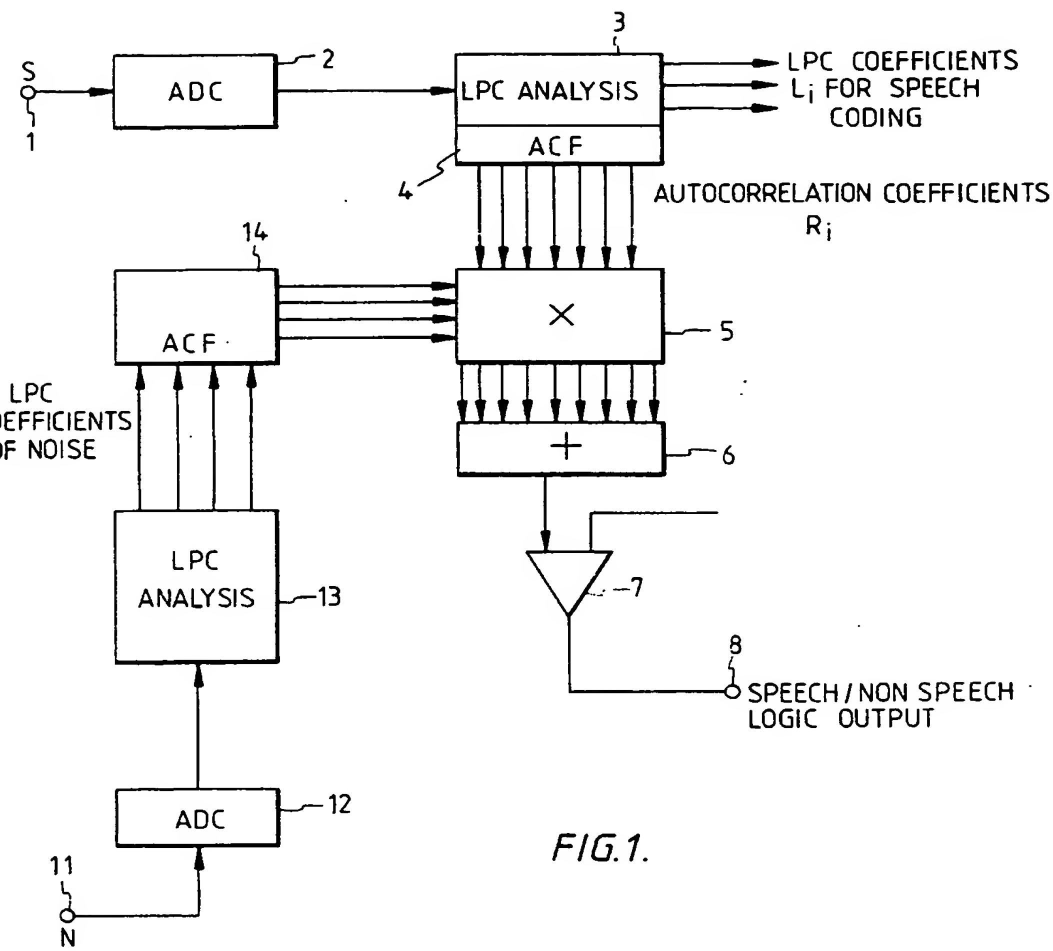

S
1
ADC
2
3
LPC ANALYSIS
ACF
4
LPC COEFFICIENTS
L_i FOR SPEECH
CODING
AUTOCORRELATION COEFFICIENTS
R_i
14
ACF
5
LPC
COEFFICIENTS
OF NOISE
6
7
LPC
ANALYSIS
13
8
SPEECH / NON SPEECH
LOGIC OUTPUT
ADC
12
11
N

FIG. 1.

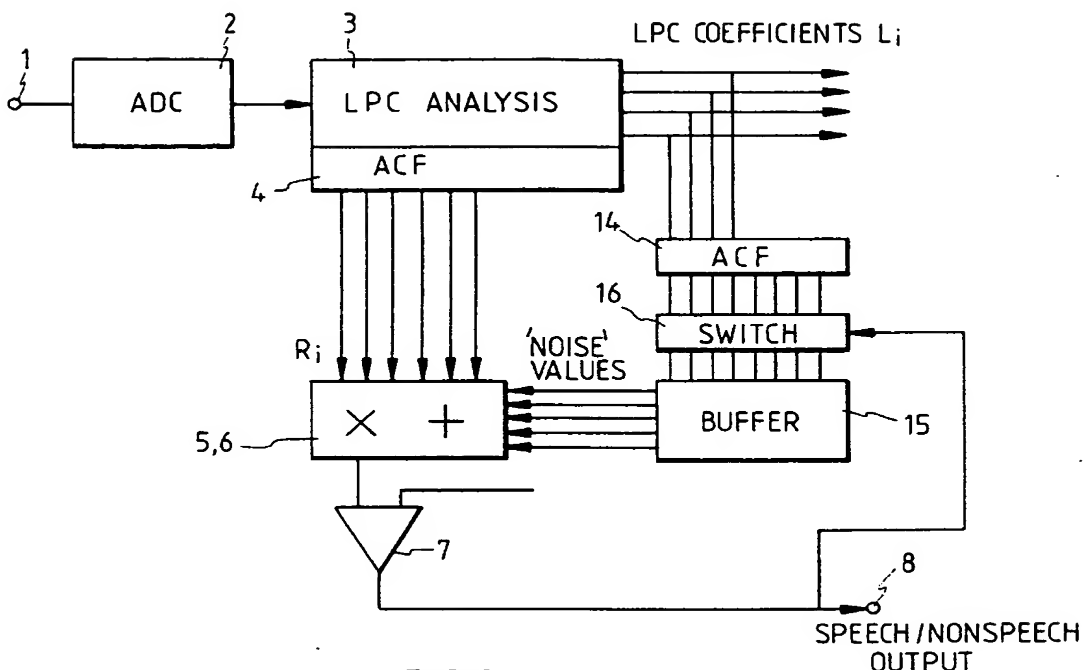
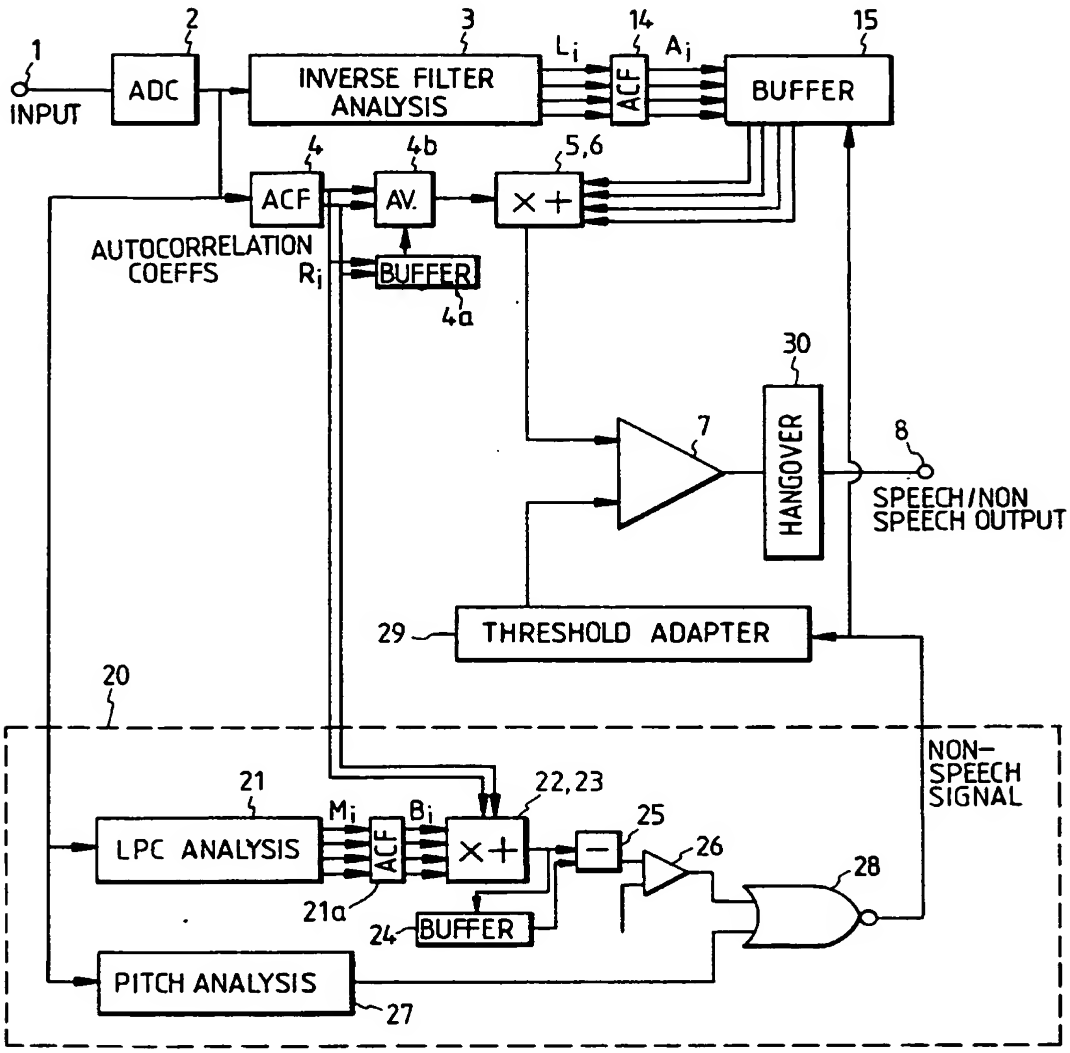

1
2
ADC
3
LPC ANALYSIS
ACF
4
LPC COEFFICIENTS Li
14
ACF
16
SWITCH
Ri
'NOISE' VALUES
BUFFER
15
5,6
7
8
SPEECH / NONSPEECH OUTPUT

FIG.2.



1
INPUT
2
ADC
3
INVERSE FILTER ANALYSIS
L_i
14
ACF
A_i
15
BUFFER
4
ACF
4b
AV.
5,6
× +
AUTOCORRELATION COEFFS
R_i
BUFFER
4a
30
HANGOVER
7
8
SPEECH/NON SPEECH OUTPUT
29
THRESHOLD ADAPTER
20
21
LPC ANALYSIS
M_i
ACF
B_i
22,23
× +
25
−
26
21a
24
BUFFER
28
NON-SPEECH SIGNAL
PITCH ANALYSIS
27

FIG. 3.